

Application of Information Theory, Lecture 6

Counting

Handout Mode

Iftach Haitner

Tel Aviv University.

November 24, 2015

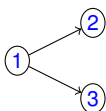
Section 1

Graph Homomorphisms

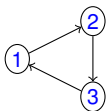
Counting # of graph homomorphisms

► $T = (V_T, E_T)$ — directed graph (no self loops)

► $G = (V_G, E_G)$



► $H = (V_H, E_H)$



► (x_1, x_2, x_3) is an **homomorphism** of G in T , if $x_1, x_2, x_3 \in V_T$ and $(i, j) \in E_G \implies (x_i, x_j) \in E_T$ (might be $x_1 = x_2$)

► Example: see board

► $\text{Hom}(X, T)$: all homomorphisms of X in T

► Claim $|\text{Hom}(H, T)| \leq |\text{Hom}(G, T)|$

► Trivial if G would be a subgraph of H

► Special case of a more general theorem

Proving the claim

- ▶ $(X_1, X_2, X_3) \leftarrow \text{Hom}(H, T)$
 - ▶ $\log |\text{Hom}(H, T)| = H(X_1, X_2, X_3)$
$$\begin{aligned} &= H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) \\ &\leq H(X_1) + H(X_2|X_1) + H(X_3|X_2) \\ &= H(X_1) + 2 \cdot H(X_2|X_1) \end{aligned}$$
(by symmetry of H)
 - ▶ Let $D_2(x)$ be the distribution of $X_2|X_1 = x$, and let $X'_2 \sim D_2(X_1)$
 - ▶
$$\begin{aligned} H(X_1, X_2, X'_2) &= H(X_1) + H(X_2|X_1) + H(X'_2|X_1, X_2) \\ &= H(X_1) + H(X_2|X_1) + H(X'_2|X_1) \\ &= H(X_1) + 2 \cdot H(X_2|X_1) \end{aligned}$$
 - ▶ $(X_1, X_2) \in E_T$ and $(X_1, X'_2) \in E_T$
- $\Rightarrow (X_1, X_2, X'_2) \in \text{Hom}(G, T)$
- $\Rightarrow H(X_1, X_2, X'_2) \leq \log |\text{Hom}(G, T)|$
- $\Rightarrow \log |\text{Hom}(H, T)| \leq \log |\text{Hom}(G, T)|. \quad \square$

Section 2

Perfect Matchings

Bregman's theorem

For bi-partite graph $G = (A, B, E)$, let $d(v) = |N(v) = \{u \in B: (v, u) \in E\}|$

Theorem 1

Let $G = (A, B, E)$ be bi-partite graph with $|A| = |B|$, and let \mathcal{M} be the perfect matchings in G . Then $|\mathcal{M}| \leq \prod_{v \in A} (d(v)!)^{1/d(v)}$.

- ▶ Let $A = B = [n] = \{1, \dots, n\}$, and for $m \in \mathcal{M}$ let $m(i)$ be the node in B matched with i by m .
- ▶ It is clear that $|\mathcal{M}| \leq \prod_{i \in [n]} d(i)$:
- ▶ Let $M \leftarrow \mathcal{M}$. Hence,

$$\begin{aligned} \log |\mathcal{M}| &= H(M) = H(M(1)) + H(M(2)|M(1)) + \dots + H(M(n)|M(1), \dots, M(n-1)) \\ &\leq H(M(1)) + H(M(2)) + \dots + H(M(n)) \\ &\leq \log d(1) + \log d(2) + \dots + \log d(n) \\ &= \log \prod_{i \in [n]} d(i) \end{aligned}$$

Proving Bregman's theorem

- ▶ Key observations:

$$H(M(i|M(1), \dots, M(i-1))) \leq \log |N(i) \setminus \{M(1), \dots, M(i-1)\}|$$

- ▶ Let \mathcal{P} be the set of all permutation over $[n]$. For $p \in \mathcal{P}$:

$$H(M) = H(M(p(1))) + \dots + H(M(p(n))|M(p(1)), \dots, M(p(n-1)))$$

- ▶ $\mathcal{S}_p(i) = \{j \in [n] : p^{-1}(j) < p^{-1}(i)\}$ — matchings proceeding i w.r.t. p

- ▶ $H(M) = \sum_{i=1}^n H(M(i)|M(\mathcal{S}_p(i)))$

- ▶ For $m \in \mathcal{M}$ and $P \leftarrow \mathcal{P}$: $|N(i) \setminus m(\mathcal{S}_P(i))|$ is **uniform** over $\{1, \dots, d(i)\}$

$$\Rightarrow \mathbb{E}_P [H(M(i) | M(\mathcal{S}_P(i)))] \leq \frac{1}{d(i)} \sum_{k=1}^{d(i)} \log k = \log ((d(i)!)^{1/d(i)})$$

$$\begin{aligned} \Rightarrow H(M) &= \mathbb{E}_P \left[\sum_{i=1}^n H(M(i)|M(\mathcal{S}_P(i))) \right] = \sum_{i=1}^n \mathbb{E}_P [H(M(i)|\mathcal{S}_P(i))] \\ &\leq \log \prod_{i \in [n]} ((d(i)!)^{1/d(i)}) . \end{aligned}$$

□

Section 3

Shearer's Lemma

$H(X_1, X_2, X_3)$ Vs. $H(X_1, X_2) + H(X_2, X_3) + H(X_3, X_1)$

- ▶ How does $H(X_1, X_2, X_3)$ compares to $H(X_1, X_2) + H(X_2, X_3) + H(X_3, X_1)$?
- ▶ If X_1, X_2, X_3 are independent, then
$$H(X_1, X_2, X_3) = \frac{1}{2} (H(X_1, X_2) + H(X_2, X_3) + H(X_3, X_1))$$
- ▶ In general: $H(X_1, X_2, X_3) \leq \frac{1}{2} (H(X_1, X_2) + H(X_2, X_3) + H(X_3, X_1))$
- ▶ A tighter bounds than $H(X_1) + H(X_2) + H(X_3)$
- ▶ Proof:

$$\begin{array}{lll} 2H(X_1, X_2, X_3) = 2H(X_1) & +2H(X_2|X_1) & +2H(X_3|X_1, X_2) \\ H(X_1, X_2) = H(X_1) & +H(X_2|X_1) & \\ H(X_2, X_3) = & +H(X_2) & +H(X_3|X_2) \\ H(X_1, X_3) = H(X_1) & & +H(X_3|X_1) \end{array}$$

- ▶ but

$$\begin{array}{l} H(X_2|X_1) \leq H(X_2) \\ H(X_3|X_1, X_2) \leq H(X_3|X_1) \\ H(X_3|X_1, X_2) \leq H(X_3|X_2) \end{array}$$

Shearer's lemma

- ▶ Let $X = (X_1, \dots, X_n)$
- ▶ For $\mathcal{S} = \{i_1, \dots, i_k\} \subseteq [n]$, let $X_{\mathcal{S}} = (X_{i_1}, \dots, X_{i_k})$
- ▶ Example: $X_{1,3} = (X_1, X_3)$

Lemma 2 (Shearer's lemma)

Let $X = (X_1, \dots, X_n)$ be a rv and let \mathcal{F} be a family of subset of $[n]$ s.t. each $i \in [n]$ appears in at least m subset of \mathcal{F} . Then $H(X) \leq \frac{1}{m} \sum_{F \in \mathcal{F}} H(X_F)$.

Proof:

- ▶ $H(X) = \sum_{i=1}^n H(X_i | \{X_\ell : \ell < i\})$
- ▶ $H(X_F) = \sum_{i \in F} H(X_i | \{X_\ell : \ell < i \wedge \ell \in F\})$

▶ Hence,

$$\begin{aligned} \sum_{F \in \mathcal{F}} H(X_F) &\geq \sum_{i=1}^n \sum_{j=1}^m H(X_i | \{X_\ell : \ell < i \wedge \ell \in F_{i,j}\}) \\ &\geq m \cdot \sum_{i=1}^n H(X_i | \{X_\ell : \ell < i\}) = m \cdot H(X) \end{aligned}$$

Corollary

Corollary 3

Let $\mathcal{F} = \{F \subseteq [n]: |F| = k\}$. Then

$$H(X) \leq \frac{n}{k} \cdot \frac{1}{\binom{n}{k}} \cdot \sum_{F \in \mathcal{F}} H(X_F) = \frac{n}{k} \cdot \mathbb{E}_{F \leftarrow \mathcal{F}} [H(X_F)].$$

Proof: $\frac{k}{n} \cdot \binom{n}{k}$ is the # of times i appears in \mathcal{F} .

Implications:

- ▶ Let $Q \subseteq \{0, 1\}^n$ and $X = (X_1, \dots, X_n) \leftarrow Q$
- ▶ $|Q| \leq 2^{\frac{n}{k} \cdot \mathbb{E}_{F \leftarrow \mathcal{F}} [H(X_F)]}$
- ▶ $\mathbb{E}_F [H(X_F)]$ is small $\implies Q$ is small
- ▶ Q is large $\implies \mathbb{E}_F [H(X_F)]$ is large

Example

- ▶ $Q \subseteq \{0, 1\}^n$ with $|Q| = 2^n/2 = 2^{n-1}$; $X \leftarrow Q$.
- ▶ $\mathcal{F} = \{F \subseteq [n]: |F| = k\}$
- ▶ By Corollary 3, $\log |Q| = n - 1 \leq \frac{n}{k} \cdot \mathbb{E}_{F \leftarrow \mathcal{F}} [H(X_F)]$

$$\implies \mathbb{E}_F [H(X_F)] \geq k(1 - \frac{1}{n}) = k - \frac{k}{n}$$

$$\implies \exists F \in \mathcal{F} \text{ s.t. } H(X_F) \geq k - \frac{k}{n}$$

- ▶ Assume $n = 1000$ and $k = 5$, hence $H(X_F) \geq 5 - \frac{1}{200}$
- ▶ X_F takes at least $2^{5 - \frac{1}{200}} = 2^{-\frac{1}{200}} \cdot 2^5 > 31$ (and hence 32) values
- ▶ Stronger conclusion: X_F is close to the uniform distribution.

More generally

- ▶ $|Q| \geq \frac{1}{2^d} \cdot 2^n$; $X \leftarrow Q$
- ▶ $\mathcal{F} = \{F \subseteq [n]: |F| = k\}$
- ▶ $n - d \leq H(X) \leq \frac{n}{k} \cdot \frac{1}{|\mathcal{F}|} \cdot \sum_{F \in \mathcal{F}} H(X_F)$

$$\Rightarrow \frac{1}{|\mathcal{F}|} \cdot \sum_{F \in \mathcal{F}} H(X_F) \geq k - \frac{dk}{n}$$

$$\Rightarrow \mathbb{E}_{F \leftarrow \mathcal{F}} [H(X_F)] \geq k - \frac{dk}{n}$$

- ▶ If $dk \ll n$, then $\exists F \in \mathcal{F}$ s.t. X_F is **close to** the uniform distribution (over k bits)

Section 4

Gold Coins

of gold coins in a cube

- ▶ Q — (finite) set of points in \mathbb{R}^3
 - ▶ Projection of Q on xy — 6
 - ▶ Projection of Q on xz — 8
 - ▶ Projection of Q on yz — 12
- ▶ Can we bound $|Q|$?
- ▶ The real story
- ▶ $X = (X_1, X_2, X_3) \leftarrow Q$
- ▶

$$\begin{aligned}\log |Q| = H(X) &\leq \frac{1}{2}(H(X_1, X_2) + H(X_1, X_3) + H(X_2, X_3)) \\ &\leq \frac{1}{2}(\log 6 + \log 8 + \log 12) \\ &\leq \frac{1}{2}(\log 6 \cdot 8 \cdot 12)\end{aligned}$$

- ▶ Hence, $|Q| \leq \sqrt{6 \cdot 8 \cdot 12} = 24$
- ▶ Can be 24 or less

of gold coins, the hyperspace case

- ▶ Q — (finite) set of points in \mathbb{R}^n
- ▶ m_i — # of coins in projection on $(1, \dots, i-1, i+1, \dots, n)$
- ▶ Claim: $|Q| \leq (\prod_{i \in [n]} m_i)^{1/(n-1)}$
- ▶ Proof: $X = (X_1, \dots, X_n) \leftarrow Q$, $X_{-i} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$
- ▶ $\log |Q| = H(X) \leq \frac{1}{n-1} \sum_i H(X_{-i}) \leq \frac{1}{n-1} \sum_i \log m_i$

Section 5

Independent Sets

of independent sets in bi-partite graphs

Theorem 4

Let $G = (A, B, E)$ be an n -regular graph with $|A| = |B| = m$. Then the number of independent sets in G is at most $(2^{n+1} - 1)m$.

Proof: \mathcal{I} — set of independent sets in G .

► Let $I \leftarrow \mathcal{I}$, let $X_v = 1$ iff $v \in I$, and $X_S = \{X_v : v \in S\}$.

► $H(I) = H(X_A | X_B) + H(X_B)$

$$\leq \sum_{v \in A} H(X_v | X_B) + \frac{1}{n} \sum_{v \in A} H(X_{N(v)}) \quad (\text{rhs by Sherer's Lemma})$$

$$\leq \sum_{v \in A} \left(H(X_v | X_{N(v)}) + \frac{1}{n} H(X_{N(v)}) \right)$$

► Fix $v \in A$. Let $\chi_v = \begin{cases} 0, & X_{N(v)} = 0^{N(v)} \\ 1, & \text{otherwise.} \end{cases}$, and $p = p(v) = \Pr[\chi_v = 0]$

► $H(X_v | X_{N(v)}) \leq H(X_v | \chi_v) \leq p$

► $H(X_{N(v)}) = H(\chi_v X_{N(v)}) = H(\chi_v) + H(X_{N(v)} | \chi_v) \leq h(p) + (1 - p) \log(2^n - 1)$

► Hence $H(I) \leq \sum_{v \in A} p(v) + \frac{1}{n} (h(p(v)) + (1 - p(v)) \log(2^n - 1))$

of independent sets in bi-partite graphs, cont.

- ▶ $\log |\mathcal{I}| = H(I) \leq \sum_{v \in A} p(v) + \frac{1}{n} (h(p(v)) + (1 - p(v)) \log(2^n - 1))$
- ▶ Let $f(t) := t + \frac{1}{n} (h(t) + (1 - t) \log(2^n - 1))$
- ▶ By calculus, $\max_{t \in [0,1]} f(t) = \frac{1}{n} \log(2^{n+1} - 1)$
- ▶ Hence, $\log |\mathcal{I}| \leq \frac{m}{n} \log(2^{n+1} - 1)$. \square

Section 6

Intersecting Graphs

Another corollary of Shearer's lemma

Corollary 5

Let \mathcal{A} and \mathcal{F} be collections of subsets of $[n]$, and for $F \in \mathcal{F}$ let \mathcal{A}_F be the collection $\{A \cap F : A \in \mathcal{A}\}$. Assume that each element of $[n]$ appears in at least m subsets of \mathcal{F} , then $|\mathcal{A}|^m \leq \prod_{F \in \mathcal{F}} |\mathcal{A}_F|$.

Proof:

- ▶ Let $X = (X_1, \dots, X_n) \leftarrow \mathcal{A}$.
- ▶ $\log |\mathcal{A}_F| \geq H(X_F)$ ($\text{Supp}(X_F) \subseteq \mathcal{A}_F$)
- ▶ By Shearer's lemma, $\log |\mathcal{A}| = H(X) \leq \frac{1}{m} \sum_{F \in \mathcal{F}} H(X_F)$. \square

of intersecting graphs

Theorem 6

Let \mathcal{G} be a family of graphs over $[n]$, s.t. $G \cap G'$ contains a triangle for each $G, G' \in \mathcal{G}$. Then $|\mathcal{G}| \leq 2^{\binom{n}{2}-2}$.

This improves over $|\mathcal{G}| \leq 2^{\binom{n}{2}-1}$, which follows from $G \cap G' \neq \emptyset$.
(wlg. all graph shares the same edge)

Proof:

- ▶ We focus on even n , and view graphs over $[n]$ as subsets $[\binom{n}{2}]$
- ▶ For $\frac{n}{2}$ -size set $\mathcal{S} \subset [n]$, let $F = F(\mathcal{S})$ be union of the cliques \mathcal{S} and $[n] \setminus \mathcal{S}$
- ▶ $F \cap G \cap G' \neq \emptyset$, for any $G, G' \in \mathcal{G}$ and \mathcal{S} as above
- ▶ Hence $|\mathcal{G}_F := \{G \cap F : G \in \mathcal{G}\}| \leq 2^{|F|-1}$
- ▶ Let $m = \binom{n}{2}$ and $m' = |F| = n(\frac{n}{2} - 1)$
- ▶ Each edge over $[n] \times [n]$, appears in $\frac{m'}{m}$ of graphs $\{F(\mathcal{S})\}_{\mathcal{S} \subset [n]: |\mathcal{S}|=\frac{n}{2}}$.
- ▶ By Corollary 5, $|\mathcal{G}|^{\frac{m'}{m} \cdot \binom{n}{2}} \leq (2^{m'-1})^{\binom{n}{2}}$
- ▶ Hence, $|\mathcal{G}| \leq 2^{m - \frac{m}{m'}} \leq 2^{\binom{n}{2}-2}$

Section 7

Statistical Distance

Statistical distance

- ▶ Let $p = (p_1, \dots, p_m)$ and $q = (q_1, \dots, q_m)$ be distributions over $[m]$
- ▶ Their **statistical distance** (also known as, variation distance) is defined by

$$\text{SD}(p, q) := \frac{1}{2} \sum_{i \in [m]} |p_i - q_i|$$

- ▶ This is simply the L_1 norm between the distribution vectors
- ▶ We will see other “distance” measures for distributions next lecture
- ▶ For $Z \sim p$ and $Y \sim q$, let $\text{SD}(X, Y) = \text{SD}(p, q)$
- ▶ Claim (HW): $\text{SD}(p, q) = \max_{S \subseteq [m]} (\sum_{i \in S} p_i - \sum_{i \in S} q_i)$
- ▶ Hence, $\text{SD}(p, q) = \max_D (\Pr_{X \sim p} [D(X) = 1] - \Pr_{X \sim q} [D(X) = 1])$
- ▶ Interpretation

Distance from the uniform distribution

- ▶ Let X be rv over $[m]$
- ▶ $H(X) \leq \log m$
- ▶ $H(X) = \log m \iff X$ is uniform over $[m]$

Theorem 7 (Next lecture)

Let X rv over $[m]$. Assume $H(X) \geq \log m - \varepsilon$, then

$$\text{SD}(X, \sim [m]) \leq \sqrt{\varepsilon \cdot \frac{\ln 2}{2}} = O(\sqrt{\varepsilon})$$